The Tao of Mathematics

Notes for a lecture at the

Indian Statistical Institute, Bangalore, 9 August, 2005.

1. Ramanujan's τ-function. I've come across your syllabus and I was impressed by the fact that Ramanujan's τ-function is taught to undergraduates. His famous work on this function dates back to 1916. As you know, the function $\tau : \mathbf{N}^{\times} \to \mathbf{Z}$ is defined formally as the coefficients in the power-series expansion of

$$\Delta(q) = q \prod_{n \geq 1} (1 - q^n)^{24}$$

$$= \sum_{n \geq 1} \tau(n) q^n,$$

where q is an indeterminate. Ramanujan made three conjectures about τ :

Conjecture I
$$\begin{cases} \tau(mn) = \tau(m)\tau(n) & \text{if } \mathrm{pgcd}(m, n) = 1, \\ \tau(p^{\alpha+2}) = \tau(p)\tau(p^{\alpha+1}) - p^{11}\tau(p^{\alpha}) & (p \text{ prime}, \ \alpha \geq 0) \end{cases}$$

Conjecture II
$$|\tau(p)| \leq 2p^{\frac{11}{2}} \quad (p \text{ prime}).$$

This is the same as saying that the reciprocals α, β of the roots of the polynomial $1 - \tau(p)\mathrm{T} + p^{11}\mathrm{T}^2$ satisfy $|\alpha| = |\beta| = p^{\frac{11}{2}}$.

Then there were many congruences (modulo 2^{11}, 3^7, 5^3, 7, 23 and 691), some of which he proved, for example

Conjecture III
$$\tau(p) \equiv 1 + p^{11} \pmod{691} \quad (p \text{ prime} \neq 691).$$

It seems that Hardy, when talking about this paper of Ramanujan, says that it belongs to the backwaters of mathematics. Weil did not fail to point out that this observation of Hardy merely shows the difference in taste between analysts and arithmeticians.

Conjecture I was proved by Mordell soon afterwards (1918). It is now a consequence of a general theory developed by Hecke. Indeed, Mordell's proof can be viewed as a precocious use of Hecke operators.

Conjecture II became part of a general theory when it was observed by Ihara that it follows from Weil's conjectures on the zeta functions of

1

varieties over finite fields, which were finally proved by Deligne in 1973. It holds the world record for the ratio

$$\frac{\text{Length of the proof}}{\text{Length of the statement}}.$$

Serre was prompted by Ramanujan's congruences to come up with his conjectures relating modular forms and representations of $\mathrm{Gal}(\bar{\mathbf{Q}}|\mathbf{Q})$; this is going to be our main concern today.

Many of the congruences in Ramanujan's conjecture III were proved by Bambah ($\tau(p) \equiv 1 + p^{11} \pmod{2^5}$, $p \neq 2$), K. G. Ramanathan ($\tau(p) \equiv 1 + p \pmod 3$, $p \neq 3$), and others.

A systematic theory was developed by Serre and Swinnerton-Dyer. They constructed, for each prime l, a representation $\rho_l : \mathrm{Gal}(\bar{\mathbf{Q}}|\mathbf{Q}) \rightarrow GL_2(\mathbf{F}_l)$ which is unramified outside l and such that for every prime $p \neq l$, one has

$$\mathrm{Tr}(\rho_l(\mathrm{Frob}_p)) \equiv \tau(p) \quad \text{and} \quad \det(\rho_l(\mathrm{Frob}_p)) \equiv p^{11} \quad (\text{in } \mathbf{F}_l),$$

where Frob_p is a lift of the automorphism $x \mapsto x^p$ of $\mathbf{F}_p$. In this way, they were able to give a uniform proof of all of Ramanujan's congruences. They could also show that no further congruences hold for the τ-function.

Now it so happens that the more fundamental function is not τ but Δ. Putting $q = e^{2i\pi z}$, we can view Δ as a function of z (in the upper half-plane $\mathrm{Im}(z) > 0$); it has some amazing properties. For example, we know that $\mathbf{SL}_2(\mathbf{Z})$ acts on the upper half-plane in a natural manner :

$$\gamma.z = \frac{az + b}{cz + d}, \qquad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{SL}_2(\mathbf{Z}).$$

It can be verified that $\Delta(\gamma.z) = (cz + d)^{12}\Delta(z)$ for every γ and every z; this is expressed by saying that Δ is a cusp form of weight 12 for $\mathbf{SL}_2(\mathbf{Z})$.

2. Modular forms. Quite generally, an analytic function f on the upper half-plane is said to be modular of weight k for a subgroup Γ of $\mathbf{SL}_2(\mathbf{Z})$ is one has

$$f(\gamma.z) = (cz + d)^k f(z), \qquad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

for every z and every $\gamma \in \Gamma$. If f vanishes at the "cusps" of Γ, it is called a cusp form. The most interesting and useful case occurs when Γ is

a congruence subgroup of $\mathbf{SL}_2(\mathbf{Z})$, i.e. defined by congruence conditions such as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod{N}$$

for some fixed integer N, which is then called the "level" of those modular forms.

Cusp forms of a given level N and weight k form a finite-dimensional vector space over $\mathbf{C}$. They have been extensively studied; tables are available online for example on William Stein's webpage. The space comes with a natural family of commuting operators, called the Hecke operators, indexed by the primes. Cusp forms which are eigenvectors for all of these operators, called eigenforms, are of special relevance; they are the ones which show up in so many places.

For example, what Wiles and his followers really proved is that every elliptic curve E over $\mathbf{Q}$, i.e. a curve of "genus 1" given by an equation of the type

$$y^2 = x^3 + ax + b \qquad (a, b \in \mathbf{Q}),$$

is "modular", i.e. there exists an eigenform f, of weight 2 and appropriate level, whose eigenvalues a_p satisfy : for almost every prime p, there are precisely $p - a_p$ solutions of the congurence $y^2 \equiv x^3 + ax + b \pmod{p}$. Gauss was the first to count the number of points on a specific elliptic curve modulo various primes; Weil the first to prove the modularity of a specific elliptic curve.

The simplest eigenform is Δ (level 1, weight $k = 12$), as was known essentially to Jacobi :

$$\Delta\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right) = (cz + d)^{12} \Delta(z), \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{SL}_2(\mathbf{Z}).$$

The eigenvalue a_p is $\tau(p)$ for every prime p. This explains why we have $11 = k - 1$ as exponent in Conjectures II and III.

We are lucky in that modular forms admit such a concrete description, as analytic functions on the upper half-plane. There is a more abstract definition, in terms of representations of the $\mathbf{GL}_2$ of the adèles of $\mathbf{Q}$, which makes it clear how they are the correct generalisation in degree 2 of characters $(\mathbf{Z}/N\mathbf{Z})^\times \to \mathbf{C}^\times$ (degree 1), and what should take their place in degree 3 etc.

Another approach to modular forms, the geometric one, sees them as sections of certain line bundles on "modular curves"; it allows us to define modular forms over arbitrary rings. Indeed, what show up in Serre's conjectures are modular forms over $\bar{\mathbf{F}}_l$.

3. Serre's conjectures. After the representation attached to Δ, Serre conjectured that there were similar representations associated to every eigenform. This was proved by Shimura in weight 2 and by Deligne in general; the weight 1 case is slightly different : in place of representations into $\mathbf{GL}_2(\bar{\mathbf{Q}}_l)$, we get (Deligne-Serre) representations into $\mathbf{GL}_2(\mathbf{C})$. Leaving aside this case, to every eigenform f, of level N and weight k, there corresponds a continuous representation $\rho_f : \mathrm{Gal}(\bar{\mathbf{Q}}|\mathbf{Q}) \to \mathbf{GL}_2(\bar{\mathbf{F}}_l)$ which is unramified outside Nl and such that for every prime p not dividing Nl, one has

$$\mathrm{Tr}(\rho_f(\mathrm{Frob}_p)) \equiv a_p \quad \text{and} \quad \det(\rho_f(\mathrm{Frob}_p)) \equiv p^{k-1} \quad (\text{in } \bar{\mathbf{F}}_l),$$

where f is sent to $a_p f$ by the pth Hecke operator. Having got this far, Serre tentatively talked about a converse.

The converse says that every odd (meaning $\det(\rho(c)) = -1$, where $c \in \mathrm{Gal}(\bar{\mathbf{Q}}|\mathbf{Q})$ is a complex conjugation) irreducible representation $\rho : \mathrm{Gal}(\bar{\mathbf{Q}}|\mathbf{Q}) \to \mathbf{GL}_2(\bar{\mathbf{F}}_l)$ arises from an eigenform by the above construction. He wrote about it to Tate in the early 70s; the reply was a proof for representations into $\mathbf{GL}_2(\mathbf{F}_2)$. Using the same technique, Serre could prove the converse for representations into $\mathbf{GL}_2(\mathbf{F}_3)$.

Some time in the mid-80s, Colmez pointed out to Serre some of the amazing consequences of his conjecture, among them the conjecture that for every elliptic curve E over $\mathbf{Q}$ is "modular", as eventually proved by Wiles and his coworkers in full generality in the late 90s. These astonishing (*inquiétantes*) consequences of his conjectures prompted Serre to make them more precise, by pinning down the level and the weight of the eigenform giving rise to the given odd irreducible representation. What is nice about this refinement is that it makes the conjecture verifiable. Given ρ, one computes the level $\mathrm{N}(\rho)$ and the weight $k(\rho)$ and checks in a finite amount of time if there is an eigenform of the type required by the conjecture.

Results of Edixhoven, Ribet and others have shown that if there is some eigenform giving rise to an odd irreducible $\rho : \mathrm{Gal}(\bar{\mathbf{Q}}|\mathbf{Q}) \to \mathbf{GL}_2(\bar{\mathbf{F}}_l)$, then there is an eigenform f of the correct weight $k(\rho)$ and the correct level $\mathrm{N}(\rho)$ giving rise to ρ. So the refined version of the conjecture follows from the naïve version.

4. Khare's proof in level 1. Following a stratagy he had worked out at the end of last year with Wintenberger, the proof is an elaborate induction on the prime l. The first few l have to be handled individually; the induction works only for sufficiently big l. It is an awsome achievement, using some of the deepest results about galoisian representations obtained

since the proof by Wiles and others that every elliptic curve over $\mathbf{Q}$ is "modular".

5. Any questions? You could ask : Why work with such a complicated field as $\mathbf{Q}$, which has infinitely many primes? Why not first study things over a simpler field such as $\mathbf{Q}_p$, which has only one prime? There are two local problems, the case $l \neq p$ and the case $l = p$. Khare himself, and Marie-France Vignéras have proved the local analogue of Serre's conjecture when $l \neq p$. Progress has been made by Breuil on the case $l = p$, where even the formulation of the problem is not easy.

You could also ask : why not first study (local and global) representations into $\mathbf{GL}_1(\bar{\mathbf{F}}_1) = \bar{\mathbf{F}}_l^{\times}$? Indeed, this is what was done from Takagi to Artin; this is now a part of the theory of abelian extensions of local or global fields. Serre's conjecture is really the first step in the programme (Langlands) of establishing a similar theory for all galoisian extensions.

6. A glimpse into the future. We began by recalling Ramanujan's three conjectures about his Δ. We have seen how all three are connected to some of the most profound mathematical theories of the last century (Hecke theory, Weil conjectures, Langlands programme).

But there is more to Ramanujan's Δ; in an article to be written by the end of this year (2005), three authors (Edixhoven, Couveignes and R. de Jong) "will clearly prove that the mod l Galois representations associated to the modular form Δ can be computed in time polynomial in l".

Chandan Singh Dalawat
Harish-Chandra Research Institute
Chhatnag Road, Jhunsi
Allahabad 211 019

Think locally

Notes for a colloquium talk at the
Indian Statistical Institute, Bangalore, 9 August, 2005.

THEOREM. — *The polynomial $x^2 + 1$ has no roots in* **Q**.

Proof : $x^2 + 1$ has no roots in **R**.

THEOREM. — *The polynomial $x^2 - 2$ has no roots in* **Q**.

The previous proof does not work, because $x^2 - 2$ does have (two) roots
in **R**. Any root in **Q** will have to belong to **Z**. For $\alpha \in$ **Z**, computing
modulo 5 gives

α	α^2	$\alpha^2 - 2$
0	0	-2
1	1	-1
2	-1	2
-2	-1	2
-1	1	-1

We see that $\alpha^2 - 2 \not\equiv 0 \pmod 5$. Hensel would have reformulated this
little computation as :

Proof : $x^2 - 2$ has no roots in $\mathbf{Q}_5$.

In general, for every prime number p, we have a locally compact field
$\mathbf{Q}_p$ which plays a similar role : it can sometimes be used to show that
something doesn't happen over **Q** because it doesn't happen over $\mathbf{Q}_p$.

The field $\mathbf{Q}_p$ can be defined in many equivalent ways. The simplest
would be to say that it is the field of fractions of the ring $\mathbf{Z}_p$, which is
integral. One could also define it as the completion of **Q** with respect to the
distance $|x - y|_p$, where the absolute value $|\ |_p$ comes from the valuation
$v_p : \mathbf{Q}^\times \to \mathbf{Z}$ sending x to the power of p in x.

Hensel was pursuing an analogy between number fields and compact
connected analytic curves. He viewed these valuations, or rather the
completions with respect to the associated absolute values, as being the
"points" on the "curve" that the number field "is". Ostrowski showed
that these are the only "points", i.e. every discrete valuation of **Q** comes
from some prime number p. This point of view is fully justified by
Grothendieck's theory of schemes, where there is indeed a "curve" (i.e. a
1-dimensional scheme) corresponding to (the ring of integers of) a number

field K whose closed points are precisely the various discrete valuations of K. Arithmetic requires that the "places at infinity", namely the completion $\mathbf{R}$ when $K = \mathbf{Q}$, be treated on an equal footing; the full significance of this requirement was realised by Arakelov.

I want to emphasize that arithmetical problems should be first studied locally, i.e. one place at a time. The trend among analysts and topologists is to seek global results; often the local versions are trivial. Not so in arithmetic. We too seek global results — applicable to number fields — but our local problems are hardly ever trivial.

As an illustration of this kind of thinking, let me recount my encounter with one your students, Anupam Kumar Singh, this morning. Along with your colleague Maneesh Thakur, he is studying "reality" of an element $x \in G(k)$ in the group of rational points of a linear algebraic group G over a field k; an element in a group is said to be "real" if it is conjugate to its inverse. They have examples of $x \in G(\mathbf{Q})$ which are not "real". My first question was to ask if that x is "real" over every place of $\mathbf{Q}$. If there is a place v where x is not "real", we have a *stronger* statement : not only is x not "real" over $\mathbf{Q}$, it is not "real" over $\mathbf{Q}_v$. If x happens to be "real" at every place, we have an instance of the failure of a *local-to-global principle*. Many people would find such an example interesting and would like to look for obstructions to account for it.

Today, we are going to work over a local field such as $\mathbf{Q}_p$. Suppose you are given a smooth projective variety over X_η over $\mathbf{Q}_p$, i.e. a variety definable by a system of homogenous polynomials

$$f_1 = 0, \ f_2 = 0, \ \ldots, \ f_r = 0 \ ; \qquad f_i \in \mathbf{Q}_p[T_0, \ldots, T_n]$$

with a condition ensuring that there are no singularities. Multiplying all the f_i by a suitable power of p, we may assume that they have coefficients in $\mathbf{Z}_p$. Reducing the f_i modulo p, we get a system

$$\bar{f}_1 = 0, \ \bar{f}_2 = 0, \ \ldots, \ \bar{f}_r = 0 \ ; \qquad f_i \in \mathbf{F}_p[T_0, \ldots, T_n]$$

defining a variety X_s over $\mathbf{F}_p$. The variety X_s need not be smooth. If it is smooth for some choice of f_i defining X_η, we say that X_η has *good reduction*; it is said to have *bad reduction* otherwise.

⟨*The rest came from the talk at Madras, at a more explicit level.*⟩

Chandan Singh Dalawat
Harish-Chandra Research Institute
Chhatnag Road, Jhunsi
Allahabad 211 019